Genomics 102 (2013) 479–483

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Complete genome sequence of the endosymbiont *Blattabacterium* from the cockroach *Nauphoeta cinerea* (Blattodea: Blaberidae)

Srinivas Kambhampati [a,*], Austin Alleman [a], Yonseong Park [b]

[a] Department of Biology, The University of Texas at Tyler, Tyler, TX 75799, USA
[b] Department of Entomology, Kansas State University, Manhattan, KS 66506, USA

ARTICLE INFO

Article history:
Received 1 May 2013
Accepted 17 September 2013
Available online 24 September 2013

Keywords:
Blattabacterium
Cockroach
Nauphoeta
Endosymbiont
Genomics
Amino acid biosynthesis

ABSTRACT

All cockroaches, with the exception of one cave-dwelling genus, harbor endosymbiotic bacteria, *Blattabacterium*. After much confusion concerning their function, recent genomic studies indicate that *Blattabacterium* synthesize amino acids, vitamins, and other compounds. However, the *Blattabacterium* genomes sequenced so far suggest that the endosymbionts are variable in their genome size, gene composition, and compounds they synthesize. Therefore, there is a need for sequencing additional *Blattabacterium* genomes to fully comprehend their evolution. Here, we report the complete genome sequence of *Blattabacterium* (BNCIN) harbored by the host *Nauphoeta cinerea* (Blaberidae). The BNCIN genome is 622,952 bp long and consists of 581 protein coding regions and 627 genes of putative function. The genome of BNCIN is comparable, with a few structural and functional differences, to the genomes of the other sequenced *Blattabacterium*. The endosymbiont is involved in complete or partial synthesis of 15 amino acids.

1. Introduction

Cockroaches (Blattodea) represent an ancient and diverse lineage of winged insects [1]. Largely tropical, a vast majority of the approximately 5000 species inhabit forested areas and generally feed on decaying, nitrogen-poor organic matter [1]. The handful of species that have adapted to an association with humans are the only ones that have been studied in detail, with little information available for the biology of those taxa that live in the wild. Thus, the biology including life history, feeding habits, and behavior remains unknown for many cockroach species.

Currently cockroaches are subdivided into five families, with some debate over whether or not the monogeneric family Cryptocercidae is monophyletic with the rest of the cockroaches [2]. The other four families are Blattidae, Blattellidae, Polyphagidae, and Blaberidae. Although there is no universal agreement on cockroach classification [3], Blaberidae is widely regarded as a derived family among cockroaches sensu stricto [1,3] with diverse biology, life history, and distribution.

All cockroaches, with the exception of *Nocticola*, the cave dwelling genus, harbor within their fat bodies endosymbiotic bacteria [4–6] of the genus *Blattabacterium*, belonging to the class Flavobacteria and the phylum Bacteroidetes [7–9]. The relationship between the host and the endosymbiont is an obligate one. Before the recent sequencing of the *Blattabacterium* genomes (see below), their function was subject to speculation [8]; however, now it is clear that the endosymbionts are involved in amino acid and vitamin synthesis from nitrogenous waste products [8,9]. Within cockroaches, surplus nitrogen, concentrated into uric acid for storage, is contained within fat body cells [10]. Cockroaches excrete ammonia as the nitrogen waste product, unlike most insects which excrete waste nitrogen as uric acid. It is likely that *Blattabacterium* are capable of utilizing both urea and ammonia because they contain an active urease as well as a urea cycle that converts host urea to ammonia [11,12]. In addition, increase in dietary nitrogen consumption by the host cockroach correlates with increased uric acid buildup within fat bodies [10,12].

Seven *Blattabacterium* genomes have been sequenced to date from the following hosts: *Periplaneta americana* (BPLAN) [13], *Blattella germanica* (BBge) [14], *Cryptocercus punctulatus* (BCpu) [15], *Blaberus giganteus* (BGIGA) [16], *Blatta orientalis* (BBor) [17], *Panesthia angustipennis* (BPAA) [18] and the termite *Mastotermes darwiniensis* (MADAR) [19]. These genomes, while largely similar to one another, also exhibit differences in structure and function, indicating considerable independent evolution among lineages harbored by various host species (see Results and discussion for details).

Blattabacterium are excellent models for studying genome evolution in endosymbionts for a number of reasons. Cockroaches represent an evolutionary lineage consisting of diverse and ancient taxa that have adapted to many habitats and exhibit broad nutritional ecology. While phylogenetic analysis supports co-cladogenesis between cockroaches and *Blattabacterium* [20,21], the gene composition among *Blattabacterium* harbored by different host species varies suggesting selection that may be correlated with host nutritional ecology. On the other hand, *Blattabacterium*, like other vertically transmitted endosymbionts, are an interesting case study in the interplay between natural selection and genetic drift because they undergo a bottleneck with each host generation.

* Corresponding author. Fax: +1 903 566 7189.
E-mail address: srini@ksu.edu (S. Kambhampati).

0888-7543/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ygeno.2013.09.003

The process by which *Blattabacterium* and other insect endosymbionts are passed on into the next generation of the host is presumably random and allows for only a subset of the bacteria to be passed on from the mother to her offspring [22]. Thus, both natural selection and random genetic drift likely play a role in genome evolution of *Blattabacterium* as was demonstrated for *Buchnera* [23] and *Blochmannia* [22].

Here, we report on the full genome sequence of *Blattabacterium* from the host cockroach, *Nauphoeta cinerea* (BNCIN), belonging to the family Blaberidae. *N. cinerea* has been the subject of many behavioral and sexual selection studies in laboratory settings [e.g., 24,25]; however, little is known about its natural history. We compare the genome structure and function of *Blattabacterium* from *N. cinerea* to those from other *Blattabacterium* and elucidate the similarities and differences among them.

2. Results and discussion

One 454 sequencing run resulted in approximately 300 Mb of DNA sequence data. The quality of the reads was excellent, with a vast majority of the reads being assigned the highest quality scores. Newbler assembled the raw reads into 56,734 contigs of varying lengths (ranging from about 100 bp to greater than 250 kb in length). Of the 56,734 contigs, 10 contigs belonged to *Blattabacterium* and the rest, presumably, to the host. The 10 contigs, made up of 155,529 reads, contained the entire genome of *Blattabacterium*. With an average read length of about 450 bp, the 155,529 reads totaled nearly 70 Mb of sequence data or an average *Blattabacterium* genome coverage of 112-fold.

The *Blattabacterium* genome (chromosome) from *N. cinerea* is 622,952 bp in length and has much in common with the other *Blattabacterium* genomes that have been sequenced to date (Table 1). All *Blattabacterium* genomes are highly reduced compared to those of their free-living relatives [26,27]. The BNCIN genome is about 98.5% identical in nucleotide sequence to that from *B. giganteus* (also in Blaberidae). The base composition was AT-biased with a GC content of 26.2% as has been the case with other *Blattabacterium* [17]. The average length of the open reading frames (ORF) is 1005 bp.

In addition, the endosymbiont harbored a 3674 bp plasmid. The bacterial annotation system, BASys [42], identified 5 genes within the plasmid. These were annotated as ribonucleoside-diphosphate reductase subunit beta, deoxyuridine 5′-triphosphate nucleotidohydrolase, and three hypothetical proteins. The number of genes in the plasmids of other *Blattabacterium* varies from 3 (BCpu) to 7 (BBor), with most containing 4 genes (Table 1). The percent identity of the BNCIN plasmid genes to plasmid genes from other *Blattabacterium* ranged from 90% to 71% with a mean of 81%. Four of the 5 genes were most similar to those contained in the plasmid of *Blattabacterium* from *P. angustipennis* [18], the host species being in the same family as *N cinerea*. The hypothetical protein 2 from the plasmid within BNCIN was most similar to that from the plasmid in BBor [17].

All existing *Blattabacterium* genomes were aligned in Geneious to examine synteny. To a large extent, the genomes were similar in structure. As noted by others [15,17,19], BCpu and MADAR are most divergent in gene composition relative to the remainder of the sequenced *Blattabacterium* genomes. The BNCIN genome, not surprisingly, was most similar in structure to that from *B. giganteus* and *P. angustipennis*, both of which are in the family Blaberidae. However, we did notice that relative to BBge, BNCIN exhibited 9 major deletions. We decided to focus on the gaps between BNCIN and BBge because the indels cumulatively spanned a few thousand base pairs and encompassed several genes. The gaps were confirmed using PCR amplification using primers anchored in the flanking regions (primer sequences available on request). Each of the gaps included one or more genes present in BBge but not in BNCIN. The genes deleted in BNCIN relative to BBge are: uracil DNA glycosylase, phosphoadenylyl-sulfate reductase, sulfate adenylyltransferase subunit 1, uroporphyrinogen-III C-methyltransferase, sulfite reductase (NADPH) hemoprotein subunits alpha and beta, hydroxymethylbilane synthase, K+ uptake transporter subunit KtrA, and hypothetical proteins BLBBGE_594 and 595. Many of the above deleted genes are involved in sulfur metabolism [28]. BNCIN is incapable of completely synthesizing methionine and retains the ability to synthesize cysteine, the two sulfur-containing amino acids. A comparison of gene composition among all genomes with the exception of BNCIN and BPAA was provided in [17].

2.1. Gene composition

We identified a total of 581 ORFs in the BNCIN genome, plus 5 more ORFs on BNCIN's plasmid. These numbers are similar to ORF counts from other *Blattabacterium* genomes sequenced to date (Table 1). Almost all the genes essential for DNA replication, RNA transcription, and mRNA translation machinery are present. Thirty-two tRNA genes, capable of transferring all amino acids and a single transfer-messenger RNA (tmRNA) gene were identified. Three ribosomal RNA genes (rRNA) were also present.

The COG composition for all sequenced *Blattabacterium* genomes is shown in Fig. 1. All the *Blattabacterium* genomes sequenced to date appear to be highly similar across the COG categories. One possible exception is COG category E (amino acid transport and metabolism), in which both BCpu and MADAR seem to have fewer genes than the other *Blattabacterium*. Not surprisingly, BNCIN, like other strains of *Blattabacterium*, has lost virtually all genes coding for cell motility, a trait not needed within the controlled environment of an insect host. COGs responsible for secondary metabolite biosynthesis, transport, and catabolism, as well as signal-transduction mechanisms are also noticeably absent within the genome.

Table 1
Genome characteristics of BNCIN compared to seven other published *Blattabacterium* strains. Host species abbreviations: BNCIN, *N. cinerea*; BGIGA, *B. giganteus*; BBge, *B. germanica*; BPLAN, *P. americana*; BCpu, *C. punctulatus*; BPAA, *Panesthia angustipennis*; MADAR, *M. darwiniensis*; BBor, *B. orientalis*. Genome size includes number of base pairs in the both the bacterial chromosome and the plasmid. The two numbers for G + C content indicate the percent G + C on the bacterial chromosome and the plasmid, respectively. Total number of genes indicates the genes in both bacterial chromosome and plasmid. The two numbers under CDS represent number for chromosome and plasmid, respectively. An asterisk indicates that the authors of the original study [18] did not provide pertinent information.

	BNCIN	BGIGA	BBge	BPLAN	BBor	BPAA	BCpu	MADAR
Genome size (bp)	622,952	632,588	640,935	640,442	638,184	632,490	609,561	590,336
Plasmids	1	1	1	1	1	*	1	1
Plasmid size (bp)	3675	3423	4085	3448	3735	*	3816	3088
Chromosome size (bp)	622,952	629,165	636,850	636,994	634,449	*	605,745	587,248
G + C content (%)	26.2/20.6	25.7/30.9	27.1/29.8	28.2/28.5	28.2/30.6	26.4/–*	23.8/30.5	27.5/31.9
Total number of genes	627	616	631	634	627	624	589	597
Total CDS	581 + 5	573 + 4	586 + 4	587 + 4	572 + 7	575*	545 + 3	544 + 4
rRNAs	3	3	3	3	3	3	3	3
tRNAs	32	34	34	33	33	34	32	34
Other RNAs: tmRNA, ffs, rnpB	1	1	3	1	3	3	3	3
Pseudogenes	5	1	1	6	9	9	3	9

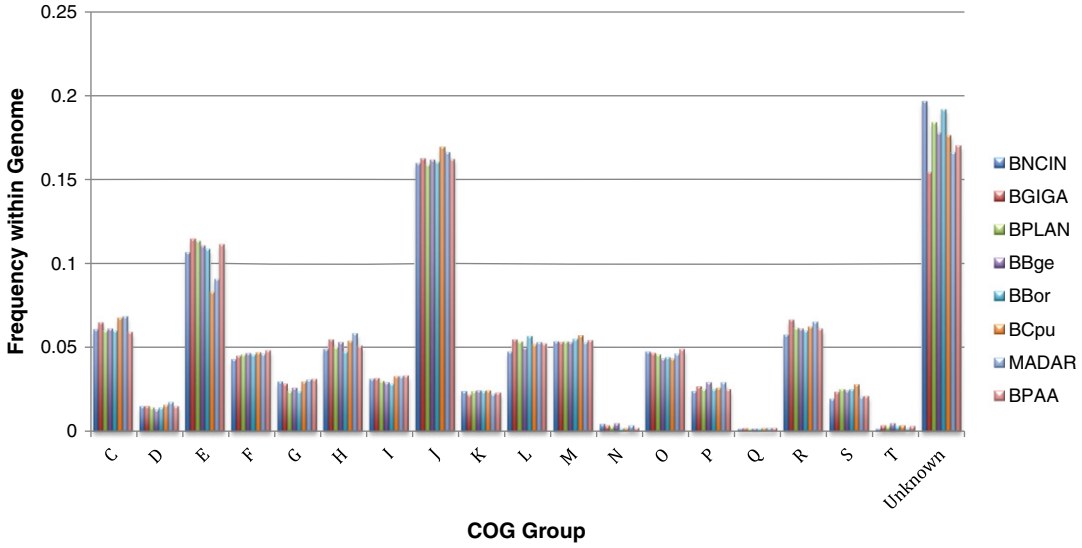

Fig. 1. COG composition of the eight *Blattabacterium* genomes. Letters refer to COG functional categories as follows. C — Energy production and conversion; D — Cell division and chromosome partitioning; E — Amino acid transport and metabolism; F — Nucleotide transport and metabolism; G — Carbohydrate transport and metabolism; H — Coenzyme metabolism; I — Lipid metabolism; J — Translation, ribosomal structure and biogenesis; K — Transcription; L — DNA replication, recombination and repair; M — Cell envelope biogenesis, outer membrane; N — Cell motility; O — Post-translational modification, protein turnover, chaperones; P — Inorganic ion transport and metabolism; Q — Secondary metabolites biosynthesis, transport, and catabolism; R — General function prediction only; S — COG of unknown function; T — Signal transduction mechanisms. Host species abbreviations are as follows: BNCIN, *N. cinerea*; BGIGA, *B. giganteus*; BBge, *B. germanica*; BPLAN, *P. americana*; BCpu, *C. punctulatus*; MADAR, *M. darwiniensis*; BBor, *B. orientalis*; BPAA, *P. angustipennis*.

2.2. Amino acid biosynthesis

We reconstructed the putative metabolic pathways using KEGG's automatic annotation server, KASS, which revealed that BNCIN has the required genes for the biosynthesis of most, but not all, amino acids. Specifically, the BNCIN genome is lacking a full complement of genes for the stand-alone synthesis of methionine, asparagine, and glutamine (Table 2). However, alternate pathways may exist for the synthesis of some of the amino acids (see below). In addition, BNCIN possesses incomplete pathways (incomplete being defined as missing the terminal step for the synthesis of a given amino acid) for alanine, glycine, proline, and serine. By comparing the amino acid synthetic pathways of BNCIN with the seven other described *Blattabacterium* genomes [13–19], it is evident that none of the *Blattabacterium* possess a full complement of genes for all of the amino acids. In general, the various *Blattabacterium* strains vary considerably in the range of amino acids they are capable of synthesizing. Only 5 amino acids, namely, histidine, phenylalanine, tyrosine, aspartate, and glutamate, are synthesized by all *Blattabacterium* strains sequenced to date (Table 2). However, excluding MADAR and BCpu, the rest of the strains are capable of synthesizing all of the essential amino acids, with the exception of methionine. Only BBge is capable of synthesizing all of the essential amino acids, including methionine [14].

Table 2
A direct comparison between seven published *Blattabacterium* genomes and BNCIN with respect to their individual abilities to synthesize the 10 essential and 10 non-essential amino acids. S represents a complete metabolic pathway, and thus the ability to synthesize the given amino acid without intervention from the host organism; – represents an incomplete metabolic pathway where the amino acid is not produced by the endosymbiont; I represents an incomplete metabolic pathway where some enzyme-coding genes, usually involved in the terminal step, are missing from the genome, but the amino acid is still synthesized by the endosymbiont. See text for details. Host species abbreviations are as follows: BNCIN, *N. cinerea*; BGIGA, *B. giganteus*; BBge, *B. germanica*; BPLAN, *P. americana*; BBor, *B. orientalis*; BPAA, *P. angustipennis*; BCpu, MADAR, *M. darwiniensis*; *C. punctulatus*.

	BNCIN	BGIGA	BBge	BPLAN	BBor	BPAA	MADAR	BCpu
Essential								
Histidine	S	S	S	S	S	S	S	S
Isoleucine	S	S	S	S	S	S	–	–
Leucine	S	S	S	S	S	S	–	–
Lysine	S	S	S	S	S	S	S	I
Methionine	–	–	S	–	–	–	–	–
Phenylalanine	S	S	S	S	S	S	S	S
Threonine	S	S	S	S	S	S	–	–
Tryptophan	S	S	S	S	S	S	–	–
Tyrosine	S	S	S	S	S	S	S	S
Valine	S	S	S	S	S	S	–	–
Nonessential								
Alanine	I	S	S	S	S	S	S	S
Arginine	S	S	S	S	S	S	S	I
Asparagine	–	–	–	–	–	–	–	–
Aspartate	S	S	S	S	S	S	S	S
Cysteine	S	S	S	S	S	S	S	–
Glutamate	S	S	S	S	S	S	S	S
Glutamine	–	–	–	–	–	–	–	S
Glycine	I	S	S	S	S	S	S	S
Proline	I	–	–	S	–	–	S	S
Serine	I	S	S	–	S	S	S	S

BNCIN's pathways for the metabolism of sulfur have been reduced when compared to the other *Blattabacterium*, although it remains more complete than the sulfur pathway possessed by BCpu, which has lost all genes needed for sulfur metabolism except for the gene encoding sulfite reductase (NADPH) flavoprotein alpha-component [15], which has been described to have additional functions such as electron relay [29] and thus may have been retained within BCpu. BNCIN has lost the ability to reduce intracellular sulfate to sulfite. When compared to BBge and MADAR, BNCIN has lost the genes required for the coding of 3′-phosphoadenosine 5′-phosphosulfate synthase and phosphoadenosine phosphosulfate reductase, although both BBge and MADAR lack the gene necessary for the conversion of adenosine-5′-phosphosulfate (APS) into 3′-phospho adenosine-5′-phosphosulfate (PAPS); BNCIN lacks the gene for this step as well. However, the BNCIN genome does encode sulfite reductase (NADPH) flavoprotein alpha-component, *cysJ*, the function of which is to facilitate the reduction of sulfate into hydrogen sulfide [29]. This sulfide is then incorporated into the two sulfur-containing amino acids L-cysteine and L-methionine. This indicates that BNCIN receives sulfate from an outside source for the production of cysteine and methionine. In BNCIN, cysteine is synthesized from L-serine via enzymes serine O-acetyltransferase and cysteine synthase A.

Like other sequenced *Blattabacterium* strains, BNCIN does not generate methionine by cysteine transsulfuration, as was previously suggested by experimental evidence [30,31]. Genomic analysis revealed

that this pathway is incomplete, lacking genes coding for homoserine O-acetyltransferase, metX, and cystathionine gamma-synthase, metB. Normally, homoserine O-acetyltransferase yields O-acetyl-L-homoserine, although in this alternate pathway, the need for acetylhomoserine is bypassed by the production of homocysteine by cystathionine beta-lyase using O-phosphohomoserine and cysteine.

While *Blattabacterium* can make arginine from glutamate, it is lacking the final step for the production of proline on its own, suggesting an incomplete ornithine–urea cycle [32,33]. There are two similar, but separate, pathways for the production of proline from the urea cycle, and genomic analysis indicated that genes for the final enzyme of both of these pathways are missing in the BNCIN genome. While the *Blattabacterium* genome does code for the ornithine–urea cycle it does not possess the gene coding for ornithine cyclodeaminase, the enzyme responsible for the conversion of ornithine into proline. Therefore, *Blattabacterium* must either acquire ornithine cyclodeaminase from its host in order to produce proline from ornithine, or it must acquire proline in its entirety from the host. Urea and ammonia are both products of proline biosynthesis through arginine degradation. These products can be used by BNCIN-encoded glutamate dehydrogenase and urease to produce glutamate. This method of proline biosynthesis does conserve nitrogen; thus, BNCIN, like other *Blattabacterium*, produces amino acids through the recycling of nitrogenous wastes.

2.3. Metabolic pathways

BNCIN has a slightly reduced capacity for DNA repair compared to BBge and BPLAN, lacking the ability to produce uracil-DNA glycosylase, which eliminates uracil from DNA molecules in order to prevent mutagenesis [34]. BNCIN also lacks coding for ATP-dependent DNA helicase (PcrA). This reduction in repair function, however, is not as severe as that displayed by MADAR, which possesses even fewer genes required for DNA repair [19].

Few substrate-specific transporters, especially those involved in amino acid uptake or secretion, were identifiable in the BNCIN genome. This is consistent with other insect nutritional endosymbionts. Like other *Blattabacterium* species, BNCIN encodes the alternate sigma factor RpoN (RNA polymerase sigma-54 factor), which functions as a transcriptional regulator of genes involved in nitrogen assimilation. While RpoN is encoded by all *Blattabacterium* sequenced to date, it is absent in most other insect sequenced endosymbiont genomes [13].

Genomic analysis of BNCIN also indicated the presence of a shortened glycolysis pathway. The genes coding for phosphofructokinase (PFK) (a metabolite needed at the beginning of the pathway to convert beta-D-fructose-6-phosphate into beta-D-fructose-1,6 biphosphate), and pyruvate kinase (PK) (required for the conversion of phosphoenolpyruvate into pyruvate) are both missing from the genome. Like other *Blattabacterium*, BNCIN likely produce pyruvate via NADP+-dependent malate dehydrogenase [14].

2.4. Conclusion

Blattabacterium within *N. cinerea* is a nutritional endosymbiont that is genetically and functionally similar to *Blattabacterium* harbored by other cockroach species, involved in nitrogen metabolism and entailing the recycling of nitrogen from ammonia and urea as well as the provisioning of amino acids to its host. *Blattabacterium* are unique among insect nutritional endosymbionts in that they have retained the transcriptional regulator RpoN, and in this respect BNCIN is similar to other *Blattabacterium*. BNCIN, like other insect endosymbionts, has a drastically reduced genome which lacks almost all specific transporters and regulatory genes, an indication of the highly specialized and dependent nature of insect endosymbionts. The metabolic capacities provided to the cockroaches by *Blattabacterium* likely played an important role in the expansion of ecological niches for cockroaches allowing for the exploitation of nitrogen-poor or nitrogen-variable food sources. Like other cockroach endosymbionts, BNCIN lacks a uricase homolog, which is integral to the utilization of uric acid as a way to store nitrogen. This step, then, is not carried out by *Blattabacterium*, but either by the cockroach host itself, or by bacteria within the insect's gut [13], as is the case with termites [35].

3. Materials and methods

3.1. Insects

N. cinerea were obtained from a colony maintained at the Insect Zoo, Kansas State University.

3.2. DNA isolation

DNA from host cockroaches was isolated as described in [13]. Briefly, freshly frozen cockroaches were dissected under a microscope to collect fat body tissue. Fat bodies were collected from 5 cockroaches directly into 200 μl of buffer ATL in the DNAeasy kit (Qiagen). The tissue was homogenized with a hand-held homogenizer and incubated at 55 °C for 1 h after the addition of 20 μl of 50 μg/ml Proteinase-K. The homogenate was filtered through a 20 μm syringe filter (Millipore). Further steps followed the protocol recommended by the manufacturer for isolating DNA from tissue.

The amount of isolated DNA was estimated on a NanoDrop 2000. To estimate the relative amounts of host and bacterial DNA, PCR was set up for two genes: a 660 bp portion of the *Blattabacterium* 16S rRNA gene using primers from Clark and Kambhampati [20] and a 500 bp portion of the insect 18S rRNA gene using primers from Kambhampati and Aldrich [36]. The PCR protocol was: 95 °C for 3 min, followed by 35 cycles of 95 °C for 30 s, 50 °C for 45 s, and 72 °C for 45 s; the final extension step was 72 °C for 10 min. The PCR products were run on a 1% agarose gel stained with ethidium bromide to visualize the DNA fragments. The relative amounts of host and bacterial DNA were estimated qualitatively.

3.3. DNA sequencing

The DNA was sent to the University of South Carolina Genomics Facility for sequencing using pyrosequencing (454 Life Sciences). The raw sequences were trimmed and assembled by Newbler software (454 Life Sciences); the assembled sequences and quality scores were electronically sent by the facility.

3.4. Genome assembly

The sequencing facility trimmed the primers, adapters, and linkers from the start or end of the sequences. Contigs were assembled using the default parameters in Newbler (454 Life Sciences). All contigs assembled by Newbler were submitted to the NCBI database using blastn and the search was restricted to *Blattabacterium*. Contigs with significant hits (e values of $\leq e^{-10}$) were identified. Each individual contig was then aligned using Clustal [37] to the full genome of *Blattabacterium* from the German cockroach, *B. germanica* [14] (GenBank Accession no. NC_013454.1, GI: 262340793) to determine the contig's position within the genome. The contigs were reverse complemented if necessary and placed in the order based on their position relative to the BBge genome.

3.5. Gene annotation

Once the entire genome was assembled relative to the reference sequence, it was submitted to GeneMark.hmm [38] for prokaryotes to identify putative ORFs. "*Blattabacterium* sp.", the only *Blattabacterium* genome available in GeneMark at the time of the analysis, was selected as the reference sequence. The putative ORFs were subjected to a blastp search to determine the extent of coverage of known BBge genes. Those predicted ORFs that consisted of only a portion of a gene (relative to the

reference genome) were manually curated to identify the missing section of that gene (i.e., we manually searched for and annotated those genes that were not completely assembled by ORF prediction software). Using a combination of manual curation and blastp searches, we assembled a predicted set of genes contained within the genome along with their physical location on the chromosome. Transfer RNA (tRNA) and tmRNA genes were identified using tRNAScan [38]. The entire plasmid sequence was contained in a single contig and was assembled and annotated as described above.

3.6. Gene ontology

To identify the putative function of each predicted gene, the curated set of genes was submitted to KEGG [39] using BBge as the reference sequence. Each identified pathway was examined for completeness and the synthesis or lack thereof of each amino acid, vitamin, and other compounds was inferred.

3.7. Comparative genomics

The assembled and annotated genome was compared to the seven published *Blattabacterium* genomes using primarily Geneious [40], Mauve [41], and NCBI's blast. Geneious was used to align the sequences and identify the indels; Geneious and Mauve were used to examine the synteny between pairs of genomes. COGs for all genomes were identified using BASys, the Bacterial Annotation System [42]. The bacterial chromosome and plasmid genome sequence have been deposited in GenBank under accession numbers CP005488 and CP005489, respectively.

Acknowledgments

We thank Mr. Sanjay Chellapilla (Kansas State University) for help with the initial bioinformatics analysis and Kiffnie Holt (Kansas State University) for cockroach specimens. We acknowledge three anonymous reviewers for constructive suggestions that helped improve the manuscript. A.A. and this research were supported by the Sam A. Lindsey Endowment to S.K.

References

[1] W.J. Bell, L.M. Roth, C.A. Nalepa, Cockroaches: Ecology, Behavior, and Natural History, Johns Hopkins Univ Press, Baltimore, 2007.
[2] D. Inward, G. Beccaloni, P. Eggleton, Death of an order: a comprehensive molecular phylogenetic study confirms that termites are eusocial cockroaches, Biol. Lett. 3 (2007) 331–335, http://dx.doi.org/10.1098/rsbl.2007.0102.
[3] M. Djernaes, K. Klaus-Dieter, M.D. Picker, J. Damgaard, Phylogeny of cockroaches (Insecta, Dictyoptera, Blattodea) with placement of aberrant taxa and exploration of out-group sampling, Syst. Entomol. 37 (2012) 65–83.
[4] P. Buchner, Endosymbiosis of Animals with Plant Microorganisms, Interscience Publishers, New York, 1965.
[5] F. Blochmann, U¨ ber das regelma¨ ssige Vorkommen von backteriena¨ hnlichen Gebilden in den Geweben und Eiern versichiedener Insekten, Z. Biol. 24 (1887) 6.
[6] M.A. Brooks, Comments on the classification of intracellular symbiotes of cockroaches and a description of the species, J. Invertebr. Pathol. 16 (1970) 249–258.
[7] C. Bandi, G. Damiani, L. Magrassi, A. Grigolo, R. Fani, et al., Flavobacteria as intracellular symbionts in cockroaches, Proc. Biol. Sci. 257 (1994) 43–48.
[8] S. Kambhampati, Family Blattabacteriaceae, 2nd ed., Bergey's Manual of Systematic Bacteriology, vol. 4, Springer, New York, 2010, p. 315.
[9] S. Kambhampati, Genus Blattabacterium, 2nd ed., Bergey's Manual of Systematic Bacteriology, vol. 4, Springer, New York, 2010, pp. 315–321.
[10] D.E. Mullins, D.G. Cochran, Nitrogen metabolism in American cockroach. I. An examination of positive nitrogen-balance with respect to uric-acid stores, Comp. Biochem. Physiol. 50 (1975) 489–500.
[11] D.E. Mullins, D.G. Cochran, A comparative study of nitrogen excretion in 23 cockroach species, Comp. Biochem. Physiol. A Comp. Physiol. 53 (1976) 393–399.
[12] M. O'Donnell, Insect Excretory Mechanisms, Advances in insect physiology, vol. 35, Academic Press, New York, 2008. 1–122.
[13] Z.L. Sabree, S. Kambhampati, N.A. Moran, Nitrogen recycling and nutritional provisioning by *Blattabacterium*, the cockroach endosymbiont, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 19521–19526.
[14] M.J. Lopez-Sanchez, et al., Evolutionary convergence and nitrogen metabolism in *Blattabacterium* strain Bge, primary endosymbiont of the cockroach *Blattella germanica*, PLoS Genet. 5 (2009) e1000721.
[15] Neef, et al., Genome economization in the endosymbiont of the wood roach *Cryptocercus punctulatus* due to drastic loss of amino acid synthesis capabilities, Genome Biol. Evol. 3 (2011) 1437–1448.
[16] C.Y. Huang, Z.L. Sabree, N.A. Moran, Genome sequence of *Blattabacterium* sp. strain BGIGA, endosymbiont of the *Blaberus giganteus* cockroach, J. Bacteriol. 194 (19) (2012) 4450–4451.
[17] R. Patino-Navarette, A. Moya, A. Latorre, J. Pereto, Comparative genomics of *Blattabacterium cuenoti*: the frozen legacy of an ancient endosymbiont genome, Genome Biol. Evol. 5 (2013) 351–361.
[18] G. Tokuda, et al., Maintenance of essential amino acid synthesis pathways in the *Blattabacterium cuenoti* symbionts of a wood-feeding cockroach, Biol. Lett. 9 (Mar. 20 2013), http://dx.doi.org/10.1098/rsbl.2012.1153.
[19] Z.L. Sabree, et al., Genome shrinkage and loss of nutrient-providing potential in the obligate symbiont of the primitive termite *Mastotermes darwiniensis*, Appl. Environ. Microbiol. 78 (2012) 204–210.
[20] J.W. Clark, S. Hossain, C.A. Burnside, S. Kambhampati, Coevolution between a cockroach and its bacterial endosymbiont: a biogeographical perspective, Proc. Biol. Sci. 268 (2001) 393–398.
[21] N. Lo, C. Bandi, H. Watanabe, C. Nalepa, T. Beninati, Evidence for cocladogenesis between diverse dictyopteran lineages and their intracellular endosymbionts, Mol. Biol. Evol. 20 (2003) 907–913.
[22] J.J. Wernegreen, Reduced selective constraint in endosymbionts: elevation in radical amino acid replacements occurs genome-wide, PLoS One 6 (12) (2011) e28905, http://dx.doi.org/10.1371/journal.pone.0028905.
[23] J.T. Herbeck, et al., A conservative test of genetic drift in the endosymbiotic bacterium *Buchnera*: slightly deleterious mutations in the chaperonin groEL, Genetics 165 (4) (2003) 1651–1660.
[24] R. Kou, et al., Juvenile hormone and the ontogeny of cockroach aggression, Horm. Behav. 56 (3) (2009) 332–338.
[25] E.L. Barrett, et al., Separate and combined effects of nutrition during juvenile and sexual development on female life-history trajectories: the thrifty phenotype in a cockroach, Proc. Biol. Sci. 22 (276(1671)) (2009) 3257–3264.
[26] M.J. Lopez-Sanchez, et al., *Blattabacteria*, the endosymbionts of cockroaches, have small genome sizes and high genome copy numbers, Environ. Microbiol. 10 (2008) 3417–3422.
[27] Z.L. Sabree, P.H. Degnan, N.A. Moran, Chromosome stability and gene loss in cockroach endosymbionts, Appl. Environ. Microbiol. 76 (2010) 4076–4079.
[28] A. Sekowska, H.F. Kung, A. Danchin, Sulfur metabolism in *Escherichia coli* and related bacteria: facts and fiction, J. Mol. Microbiol. Biotechnol. 2 (2000) 145–177.
[29] L.M. Siegel, P.S. Davis, Reduced nicotinamide adenine dinucleotide phosphate-sulfite reductase of enterobacteria. IV. The *Escherichia coli* hemoflavoprotein: subunit structure and dissociation into hemoprotein and flavoprotein components, J. Biol. Chem. 249 (1974) 1587–1598.
[30] R.J. Block, S.M. Henry, Metabolism of sulphur amino-acids and of sulphate in *Blattella germanica*, Nature 191 (1961) 392–393.
[31] S.M. Henry, R.J. Block, Sulfur metabolism of insects. 6. Metabolism of sulfur amino acids and related compounds in German cockroach, *Blattella germanica* (L.), Contrib. Boyce Thompson Inst. 21 (1961) 129–145.
[32] T. Inokuchi, Y. Horie, T. Ito, Urea cycle in silkworm *Bombyx mori*, Biochem. Biophys. Res. Commun. 35 (1969) 783–787.
[33] S.R. Reddy, J.W. Campbell, Enzymic basis for the nutritional requirement of arginine in insects, Experientia 33 (1977) 160–161.
[34] T. Lindahl, S. Ljungquist, W. Siegert, B. Nyberg, B. Sperens, DNA N-glycosidases: properties of uracil-DNA glycosidase from *Escherichia coli*, J. Biol. Chem. 252 (1977) 3286–3294.
[35] G. Tokuda, H. Watanabe, Hidden cellulases in termites: revision of an old hypothesis, Biol. Lett. 3 (2007) 336–339.
[36] S. Kambhampati, B.T. Aldrich, Inexpensive molecular weight markers, J. Entomol. Sci. 41 (2006) 412–415.
[37] M.A. Larkin, G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez, J.D. Thompson, T.J. Gibson, D.G. Higgins, Clustal W and Clustal X version 2.0, Bioinformatics 23 (2007) 2947–2948.
[38] A. Lukashin, M. Borodovsky, GeneMark.hmm: new solutions for gene finding, NAR 26 (4) (1998) 1107–1115.
[39] M. Kanehisa, S. Goto, Y. Sato, M. Furumichi, M. Tanabe, KEGG for integration and interpretation of large-scale molecular datasets, Nucleic Acids Res. 40 (2012) D109–D114.
[40] A.J. Drummond, B. Ashton, S. Buxton, M. Cheung, A. Cooper, C. Duran, M. Field, J. Heled, M. Kearse, S. Markowitz, R. Moir, S. Stones-Havas, S. Sturrock, T. Thierer, A. Wilson, Geneious v5.4, Available from http://www.geneious.com/ 2011.
[41] A.C.E. Darling, B. Mau, F.B. Blatter, N.T. Perna, Mauve: multiple alignment of conserved genomic sequence with rearrangements, Genome Res. 14 (7) (2004) 1394–1403.
[42] G.H. Van Domselaar, P. Stothard, S. Shrivastava, J.A. Cruz, A. Guo, X. Dong, P. Lu, D. Szafron, R. Greiner, D.S. Wishart, BASys: a web server for automated bacterial genome annotation, Nucleic Acids Res. 33 (Web Server Issue) (2005) W455–W459.